

Systematic Literature Review of Machine Learning Methods in Insulin Secretion Model Analysis

Mohd Hussaini Abbas, Nor Azlan Othman*, Samsul Setumin, Nor Salwa Damanhuri, Rohaiza Baharudin, Nur Sa'adah Muhamad Sauki and Sarah Addyani Shamsuddin

Abstract— Endogenous insulin secretion (U_N) plays a critical role in maintaining glucose homeostasis. Pathological changes in U_N enable early detection of metabolic inefficiency prior to the onset of diabetes mellitus (DM). Numerous researches have been carried out to establish the most effective method for assessing the participant's glycemic state by identifying their U_N profile. In contrast to insulin sensitivity (SI), there is no gold standard for U_N profile. Thus, the deconvolution of C-peptide measurements is used in the majority of research to identify the U_N profile. Due to the fact that C-peptide and insulin are co-secreted equimolarly from pancreatic β -cells, the latter method is shown to be accurate. Although studies have shown that the machine learning-based strategies can yield very positive outcomes in other areas of DM diagnosis, there is currently little research that employing machine learning for quantifying the U_N profile to enable early diagnosis of metabolic dysfunction. Hence, the main objective of this study is to conduct a thorough search on machine learning-based modelling strategies that were used to identify the individual-specific U_N profile through the development of a U_N model. Additionally, this study will investigate whether the data acquired from the U_N model can be used to quantify a person's metabolic condition (either normal, pre-diabetic or T2D). The literature search turned up prospective studies linking machine learning and U_N in its search and analysis. Meta-analyses summarize the available data and highlight various methodological stances. Thus, the exploratory of machine learning classification and regression technique can be portrayed in 3 different scenarios during the identification of U_N profile. The 3 scenarios are: the study of insulin secretion through analyzing the insulin sensitivity, the study of U_N without taking into considerations or in-depth study of U_1 and U_2 , and the study of insulin secretion using deconvolution of plasma C-peptide concentrations. It is evident that while Decision Tree (DT) is ideal for the first scenario, Random Forest (RF) is the better option for the other two scenarios. Further optimization can be implemented with the use of these techniques under supervised learning to improve diagnosis and comprehend the pathogenesis of diabetes, particularly in U_N .

Index Terms—diabetes, insulin secretion, machine learning.

This manuscript is submitted on 23rd Feb 2023 and accepted on 24th July 2023. Mohd Hussaini Abbas, Nor Azlan Othman, Samsul Setumin, Nor Salwa Damanhuri, Rohaiza Baharudin, Nur Sa'adah Muhamad Sauki and Sarah Addyani Shamsuddin are with Electrical Engineering Studies, College of Engineering, Universiti Teknologi MARA Cawangan Pulau Pinang, Kampus Permatang Pauh, 13500 Permatang Pauh, Pulau Pinang, Malaysia

*Corresponding author
Email address: azlan253@uitm.edu.my

1985-5389/© 2023 The Authors. Published by UiTM Press. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

I. INTRODUCTION

DIABETES Mellitus disease is becoming more prevalent around the world, with the number of cases increasing at an alarming rate. Diabetes now affects over 420 million people globally, making it one of the leading causes of death [1]. In Malaysia, citizen habits of consuming a high-sugar diet have put them at risk of diabetes as early as the age of 18 [2].

Commonly, Malaysian consumes lots of rice-based meal throughout their daily diet. Studies show that high consumptions of rice-based meals, especially white rice, leads to a high intake of carbohydrates. The human digestive system subsequently breaks down this carbohydrate into sugar, resulting in a rise in blood glucose (BG) levels [3-5]. Furthermore, not only the rice-based meal, the increased level of BG might be effected by exacerbated of drinking high-sugar beverages [6].

Diabetes is commonly related to the inability of the human body to reach glucose homeostasis. In order to reach a glucose homeostasis state, insulin is required to initiate the glucose uptake process. As a level of glucose in blood stream rises, the pancreas senses this increase and stimulates to release insulin into blood stream. This newly released insulin plays an important role in regulating the concentration of glucose in blood [7]. Generally, a healthy fasting BG level is in the range of 4 – 5.6 mmol·L⁻¹. Failing to continually control the BG levels leads to the development of metabolic disorders, particularly diabetes, all of which have significant complications that can reduce quality and length of life.

A. Diabetes

Diabetes is a chronic condition that arises when the pancreas fails to produce enough insulin or when the produced insulin is ineffectively utilized. Critically, There are two major types of diabetes which are Type 1 and Type 2 diabetes [8, 9]. Type 1 diabetes (T1D) results from the body's failure or inability to produce insulin. The etiology of T1D is characterized by the destruction of the pancreatic β -cells caused by an auto-immune disorder or as a result of the action of genetic markers [9]. Individuals with T1D need to have an insulin injection in order to maintain a safe level of glucose in the blood stream.

Type 2 diabetes (T2D), on the other hand, results from the body's inability to make enough (insulin deficiency) or properly use endogenous insulin secreted by the β cell of the pancreas (insulin resistance). Prior studies have characterized T2D by fasting hyperglycemia and an excessive rise in the BG

level above baseline following post-prandial [9]. Thus, it can be clearly seen that understanding the physiological characteristic of insulin produced by pancreatic β -cells could potentially ameliorate the worst symptoms of this diabetes.

B. Insulin Secretion

In order to reach a glucose homeostasis, sufficient insulin production is needed to be secreted internally and efficiently. Insulin secretion, U_N is generally produced by pancreatic β -cells in response to a quick or steady rise in BG levels. The glucose level, which must be lowered down to the normal range of 4–5.6 mmol·L⁻¹, determines the amount of insulin produced [9, 10]. Mathematical deconvolution is often used to model insulin production as a function of peripheral C-peptide levels. Prior studies have adopted the endogenous U_N profile confirmed by Van Cauter's model due to the fact that C-peptide and insulin are co-secreted by β -cells [9, 11]. However, C-peptide assays are considerably scarce, costly and time consuming, even though they are still the best approach for diagnosing insulin secretion, U_N .

While C-peptide testing is diagnostically useful, there is still room to improve U_N profile identification without using the C-peptide profile. However, detecting U_N purely based on C-peptide measurements will be hampered since the pharmacodynamic reactivity of β -cells to glucose is not taken into consideration [12-28]. Furthermore, endogenous insulin's regulation of BG is basically a closed-loop feedback-control mechanism, with secretion responding to glucose level and rate of change. Thus, by using machine learning technique, hypothetically there is a possibility to identify the individual specific U_N profile without the use of C-peptide.

C. Machine Learning

Recent studies have been done to evaluate the performance of U_N towards glucose homeostasis engaging the use of machine learning and classical statistical modeling [29-34]. Additionally, a study shows that neural network has the best performance in T1D and T2D classification problems [23, 29, 35]. Understanding the individual-specific behavior of U_N profile is very crucial in distinguishing the metabolic state of the individual; either normal, pre-diabetic or T2D.

There are numerous types of approach that are being utilized, each of which provides varied levels of accuracy and appropriateness for the data acquired [12-25, 27, 28, 36-43]. Although machine learning-based approaches have shown to produce promising results in other areas of diabetes mellitus diagnosis, there is currently little or no research that explores the use of machine learning-based modelling approaches for quantifying the individual-specific U_N profile to enable early metabolic dysfunction diagnosis. In addition, prior studies show that U_N is secreted in a bi-phasic manner; first phase secretion (U_1) and second phase secretion (U_2). Further linkage between anatomical function (age, sex and body mass index) towards the secretion of U_1 and U_2 is remained unclear. No effort has been made so far to develop a machine learning-based approach for formulating controller gains of U_1 and U_2 in the context of individual-specific anatomical function.

Hence, there is a need to explore machine learning approach in analyzing the individual-specific U_N model. Thus, by using a systematic literature review (SLR) technique, this study will focus on reviewing the latest and best methods that have been applied by prior studies to model this individual-specific U_N profile using machine learning approach. The outcome of this study will potentially provide a better solution for future research in identifying the individual-specific U_N profile.

II. METHODOLOGY OF SYSTEMATIC LITERATURE REVIEW (SLR)

This study focuses on identifying, evaluating, categorizing, analyzing and interpreting all available research on U_N model analysis engaging the use of machine learning approach. A thorough literature search was conducted to identify systematic review papers for this study. Descriptions of these processes are summarized in several topics in the next few paragraphs.

A. Research Objective

The main objective of this study is to search on the U_N models that were developed between 2016 and 2021. A thorough search on machine learning-based modeling strategy that were used to identify the individual-specific U_N profile through the development of a U_N model rather than the traditional approach of deconvolution of C-peptide measurements to identify the U_N profile will be presented here. Furthermore, this study will examine how the data acquired from the U_N model might be useful to distinguish an individual's metabolic condition (either normal, pre-diabetic or T2D).

B. Search Process

An automatic search in five bibliographic databases as well as a human (manual) search in the list of references of current secondary research in the field was established as search tactics. To aid this research, both paid and open-source platforms were utilized. The Tun Abdul Razak Library, PTAR UiTM, subscribes to and provides this service to students, so they can access the paid platform. While open source is gathered through websites such as ScienceOpen, Semantic Scholar, CORE, and others that allow free access to scientific papers.

1) Automated Search

As an open-source platform, Google Scholar makes it easy to conduct a broad search for scholarly literature. This platform facilitates and is very helpful in obtaining official reports, journals and conference proceedings along with other information in related fields of technology and science. With capabilities such as the ability to search all scholarly literature in one location, Google Scholar can assist in identifying relevant research across the world of academic publishing.

Several major databases have been selected for obtaining relevant journals. The IEEE Xplore, Science Direct, Scopus, SpringerLink and the Web of Science were chosen as well-known databases that index important journals and conference proceedings in the fields of technology and computer sciences, as well as based on the authors' analysis of preliminary investigation results. Scopus and the Web of Science (WoS) are fundamentally different in that they both index studies from a variety of sources based on particular criteria. As a result, while

identical documents may be obtained from multiple databases, each database is anticipated to contribute a number of unique documents, ensuring that variety is ensured. The search was conducted in September 2021 and was limited to papers published between 2016 and 2021, with English as the only language.

The following generic search string was developed based on the keywords discovered in the research objective and their synonyms. At first, the generic search string used was “TITLE: (diabetes) AND (machine learning)”. However, the total number of papers related to previous search string was huge, approximately 6183 papers where 1927 papers was from WoS, 2744 papers from Scopus, 890 papers from IEEE, 622 papers from Science Direct and zero paper from SpringerLink. Further refinement had been made by changing the generic search string to “TITLE: (insulin secretion) AND (machine learning)” resulting in 95 papers where 26 papers was from WoS, 26 papers from Scopus, 2 papers from IEEE, 41 papers from Science Direct and zero paper from SpringerLink. Table I tabulates the number of papers gathered based on difference generic search string. By using search string to TITLE: (insulin secretion) AND (machine learning), the scope of the search can be narrowed, and topic specialization can be implemented.

TABLE I
THE NUMBER OF PAPERS GATHERED BASED ON DIFFERENCE GENERIC SEARCH STRING

Year	E-Resources	Generic Search String	
		Diabetes, Machine learning	Insulin secretion, Machine learning
2016	WoS	1927	26
	Scopus	2744	27
-	IEEE	890	2
2021	Science Direct	622	42
	SpringerLink	0	0
Total number of papers		6183	97

2) Manual Search

The manual search from primary studies (97 papers gathered through automated search) was carried out by looking through the references lists of each study on the topics of “insulin secretion” and “machine learning” using Google Scholar platform. The references lists were randomly checked for titles then compared to the reference list produced from the automated search approach. This procedure is carried out in order to identify papers that may have been overlooked or are not listed in any of the five bibliographic databases.

C. Selection Criteria

Following the automated search, a three-phase manual selection of relevant documents is carried out: screening, eligibility, and inclusion. The goal of the screening step is to limit the amount of documents that must be reviewed in their entirety by evaluating only specified parts of all documents first, and then reading the complete text of the accepted documents. The papers' full texts have been downloaded and examined. Each abstract will be evaluated to see whether the paper is related to engineering or focuses only on medical issues. Preferably, the selected journal uses a machine learning-

based approach. Two sets of selection criteria were defined: inclusion and exclusion criteria, which are described in Table II. These criteria are being set in order to eliminate bias during the three-phase process.

There are only two inclusion criteria, which fundamentally divide original works into two categories: works that proposed a model or approach to classify the subject is normal, pre-diabetic or diabetic-type2 were accepted, as well as works that presents one or more methods using machine learning. Papers related to T1D and Gestational diabetes mellitus (GDM) were removed as it was outside the area of specialization of this study. Additional exclusion criteria can be used to clarify the reasons for exclusion since a document under evaluation can have numerous selection criteria. Some of these factors deserve special attention.

TABLE II
THE INCLUSION AND EXCLUSION CRITERIA

Inclusion criteria	
IC01	The study presents a model or approach to identify U_N .
IC02	The study presents a model or approach to classify the subject either normal, pre-diabetic or diabetic-type2.
IC03	The study presents one or more methods using machine learning.
Exclusion criteria	
EC01	The study is not available in full-text.
EC02	The study is not available in English.
EC03	The study is a duplicate.
EC04	The study was not published between 2016 and 2021.
EC05	The study too focused on the medical aspect.
EC06	The study does not mention the dataset utilized to validate its method.
EC07	The study does not present a model or approach to identify U_N .
EC08	The study does not present a model or approach to classify the subject is normal, pre-diabetic or diabetic-type2.
EC09	The study does not present one or more methods using machine learning.

As in EC01, some documents are not fully accessible especially when the search is done through the Google Scholar platform. In addition, this platform only displays the abstract or a brief summary of the paper. For EC02 criteria, there are some documents that use languages other than English. Documents employing Malay, Mandarin and Tamil languages are among those found. Such documents had to be rejected for fear of not being able to interpret the true meaning that the author of the paper wanted to convey. When publications are published in two or more databases, the EC03 criterion applies. Similar works can be found, for example, in the IEEE and Scopus databases. The goal of the EC04 criterion is to acquire the most recent articles and research over the last five years. This is critical in ensuring that the literature review is up to date with the most recent findings and applications in perspective of science's rapid advancement.

The EC05 criteria were intended to eliminate documents that go into too much detail about medical issues. This is due to the fact that the investigation will be undertaken from an engineering point of view, especially when it comes to modelling and using machine learning as a classification approach. When a study fails to mention the origin of the dataset used to validate its approach, it fails to meet criterion EC06. This is to ensure that the data obtained is reliable and accurate,

particularly when it comes to medical data (clinical ethics approval has to be enclosed). Although the main objective of this systematic literature review (SLR) paper is to focus on engineering, the outcomes must be valid and portray a convincing analysis from a medical perspective.

The EC07, EC08 and EC09 criteria were supposed to eliminate articles that did not present a model or approach for identifying U_N profile, distinguishing the metabolic state of the individual either normal, pre-diabetic, or T2D, or that did not present one or more machine learning methods. If no exclusion criteria are met, a document is approved. An exclusion criterion, on the other hand, is only assigned to a document if it can be verified: if in doubt, a work is always accepted and subject to a more thorough evaluation in the next selection phase.

D. Data Extraction Form

A data extraction form was developed to be applied to each study included in the review after the selection processes in order to collect evidence to answer the research objectives presented in Section II - A. The extraction proposed to accommodate data acquired from the investigations was likewise guided by the research topics. (1) Metadata, (2) Patient, (3) Problem Statements, (4) Methodology, (5) Modelling Approach, (6) Results, (7) Weakness and (8) Research Gap are the eight sections of the form. Table III provides a thorough overview of the fields in the data extraction form. These data are extracted and gathered from the reading of each journal publication obtained through Section II -D.

TABLE III
THE DATA EXTRACTION FORM

Extracted Data	Description
(1) Metadata	Title, authors and publication year of the study.
(2) Patient	The subject in the study either human or animal (rat or mice).
(3) Problem Statements	The problem statement occurs until the study was conducted.
(4) Methodology	The methodology used in the study to gather and analyze dataset from the related samples (clinical trial).
(5) Modelling approach	The modelling approach used to identify the U_N profile and distinguish the metabolic state of an individual.
(6) Result	The highlighted advantages from the result and analysis of study.
(7) Weakness	The highlighted disadvantages from the result and analysis of study.
(8) Research Gap	The point of view that can be considered for improvement in the study.

E. Search and Selection Results

A total of 97 articles were obtained through the methods discussed in Section II -C. Because there are crosses of documents indexed in each database as well as listed as paper references picked as manual search sources, duplicates are expected to be retrieved. After eliminating duplicates, a total of 95 unique articles remained, resulting in the initial collection of candidate documents to be reviewed in this study.

Only the title, keywords, and abstracts of these 95 unique articles were reviewed during the first phase of preliminary document selection process. This activity included surface reading, which assessed each document's suitability for the

previously specified inclusion and exclusion criteria. As a result, in this round, 84 articles were accepted, that is, at a rate of 88.42%. At this phase, 11 articles were rejected (11.57%).

During the second phase (final selection), a thorough reading of the full-text manuscript was carried out to further examine the suitability of these 84 articles towards the objective of this SLR paper. As a result, from final selection, 30 articles were preserved in this process, accounting for 35.71% of the reviewed documents in this phase. The search and selection techniques are summarized in Fig 1.

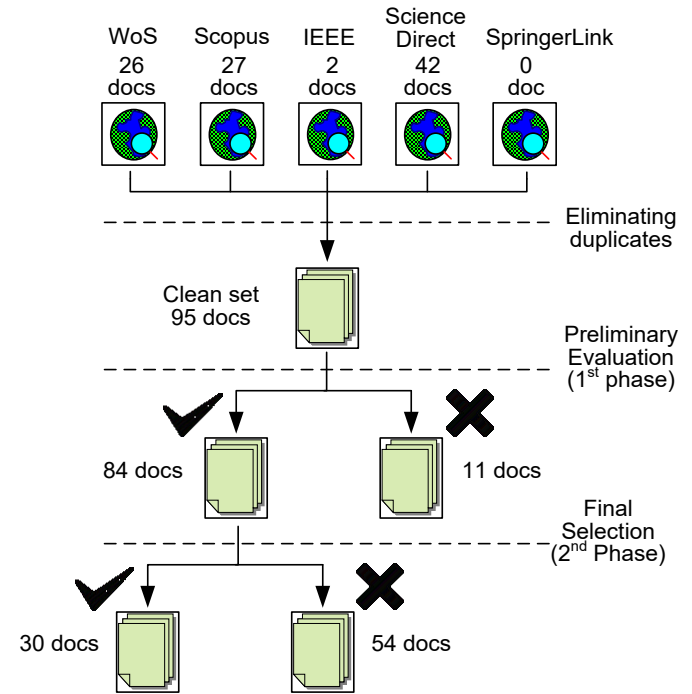


Fig. 1. Flow of search and selection procedures.

III. RESULTS AND DISCUSSION

In order to answer the research objective for this SLR as in Section II - B, several factors must be considered. This will also serve as a guide to take the study to a more extensive and in-depth level. This section will go through the results from Section II and how they were analyzed. The patient category, the research gap, and machine learning regression and classification are among the criteria that looked into. One of the top machines learning analytical techniques will be identified and highlighted for this study at the end of the discussion.

A. Patients

In these 30 articles, it can be seen that most studies used humans or animals as patients' data. More succinctly, one research utilized mice as patients, one study employed both rats and humans as patients, and the remaining 28 studies used humans as patients' data. The use of animals was intended to decrease risk and can be considered as a preliminary study before implementing it to humans for further research. The use of animal is only 6.67% of the total number of studies available. Although the percentage is low, however the outcomes from the research is still valid as it is presumed that the physiology of the chosen animals (rat or mice) is closed to humans [44]. Because

of the differences between mice and human, it is possible that medications for diabetes reversal in rat or mice have not been successfully translated to humans. Increased understanding of the human intricacies could lead to novel approaches to reversing diabetes in humans [45].

B. Research Gap

The research gap for this work is grouped into many categories using the procedure described in Section II -E. The research gap statement and percentage for the available studies are shown in Fig 2. This Fig. 2 portrays the result extracted from 30 studies that mainly focus on machine learning approach towards insulin secretion. It can be clearly seen that 73.33% of the studies focus on analyzing the insulin sensitivity (SI) to identify the participant-specific U_N profile. Hypothetically, higher value in SI would indicate a good production of U_N . In addition, if the SI value is low, it will give the impression that the receptor of muscle and adipose tissue cells is unable to bind with insulin to mediate the glucose uptake. However, it can be safely said that the amount of U_N produced is higher than normal production (hyper insulinaemic condition). Based on the analyzed SI values, these studies classify them into each metabolic state, either normal, pre-diabetic and T2D. However, it is understood that solely relying on SI value to determine the metabolic state of the participant is considered to be less convincing. Hypothetically, it is best to have a detailed analysis of both SI and U_N profile in order to distinguish the metabolic state of the participant. Studies show that the etiology of T2D comes from insulin resistance ($IR = SI^{-1}$) and insulin deficiency [9, 46].

On the other hand, 16.67% of the studies apply machine learning approach in identifying the participant-specific U_N profile only without considering the first (U_1) and second (U_2) phase of the U_N profile. Theoretically, U_N is secreted in a biphasic manner [47]. Hypothetically, by understanding the physiological pattern of these U_1 and U_2 profiles, among the classifiers that exist and are often used in studies over the past five years, ameliorate the analysis of the pancreatic β -cell function [9]. In addition, the final 10% of the studies also employ the machine learning approach to identify the participant-specific U_N profile without considering these two phases. However, the only difference between 16.67% and 10% lies in the data collection. 16.67% uses glucose measurements to identify the U_N profile whereas 10% uses C-peptide measurements to identify the U_N profile. Although studies show that identifying the U_N profile using C-peptide measurements prove to be accurate due to the fact that U_N is equimolarly secreted with C-peptide, the C-peptide measurements are relatively sparse, expensive and time consuming [48, 49]. As diagnostically effective, there remains scope to better identify U_N profile without using C-peptide measurement.

C. Machine Learning Regression and Classification

Hypothetically, there are 2 stages under machine learning technique when analyzing the U_N profile. These stages are regression and classification.

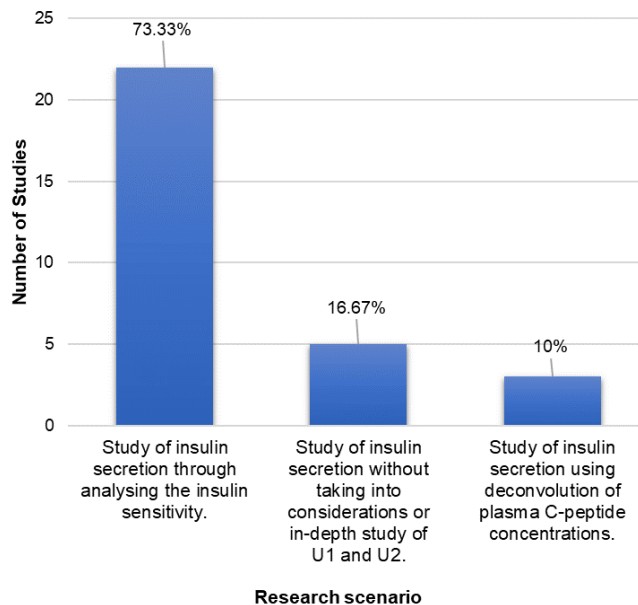


Fig. 2. Three Different Scenarios in Identifying the U_N Profile Using Machine Learning.

1) Regression

A supervised learning process called regression is used in machine learning to forecast the continuous output value of a dependent variable based on the input values of independent variables. The purpose of regression is to create a model that can predict the values of one or more independent variables. Regression algorithms generate a model from a training set of data that can subsequently be used to predict the target variable in fresh or unknown data. The model is assessed based on its ability to predict the target variable in the test set accurately. The ultimate purpose of regression is to create a model that can forecast the target variable reliably for fresh data.

Regression is a statistical approach used to study the connection between one or more independent variables and a dependent variable. In the context of identifying a U_N profile utilizing C-peptide or glucose profile, regression analysis may be utilized to establish the strength of the association between these variables and the U_N profile. C-peptide is a chemical generated by the pancreas during insulin processing. Its blood levels can be utilized as an indication of insulin production. The glucose profile, on the other hand, is a measurement of blood sugar levels across time. C-peptide and glucose profile are both glucose metabolism markers that are extensively utilized in the diagnosis and management of diabetes.

A dataset containing measurements of C-peptide or glucose profile, as well as U_N profile, would be necessary to utilize regression to determine a U_N profile. After that, the dataset may be analyzed using linear regression, which includes fitting a line to the data points to reflect the connection between the independent variable(s) (C-peptide or glucose profile) and the dependent variable (U_N profile). The Pearson correlation coefficient or Spearman's rank correlation coefficient can be used to quantify the strength of the link between the C-peptide or glucose profile and the U_N profile [50]. These coefficients vary from -1 to 1, with 1 indicating perfect positive correlation, -1 indicating perfect negative correlation, and 0 indicating no

association. After determining the association between C-peptide or glucose profile and U_N profile, the regression model may be used to predict U_N profile based on C-peptide or glucose profile data. This data may be utilized to diagnose and manage illnesses like diabetes, as well as to optimize treatment approaches.

In general, in order to identify the participant-specific insulin secretion, blood samples are required. These blood samples are then assayed for glucose, insulin and C-peptide measurements using standard commercial assays at an accredited laboratory. Although it is claimed that the use of machine learning is good enough, using regression of machine learning with C-peptide data will result in debatable results due to C-peptide issue (sparse measurement). Thus, it can be clearly seen that in Fig. 2, less research (10%) has been done in using regression of machine learning to quantify insulin secretion using deconvolution of plasma C-peptide concentrations. Prior research uses Random Forest (RF) as the preferred technique on identifying U_N profile [12, 22, 26, 27, 41]. With a smaller number of researches being done to identify U_N profile using the regression of machine learning, there is a need to investigate, in depth, on how to manipulate the regression technique in accurately identify the U_N profile. Hypothetically, there is a possibility to use regression technique with glucose measurement when quantifying the U_N profile. Study also shows that glucose measurement profile mimics the behavior of U_N profile in terms of U_1 and U_2 profile [40, 42, 43, 51, 52]. Hence, it will deliver potential real diagnostic and monitoring metrics.

2) Classification

Classification is a form of supervised learning method used in machine learning to predict a dependent variable's category output value based on the input values of independent variables. The dependent variable in classification is either categorical or discrete, and the objective is to develop a model that can reliably predict the category of the dependent variable based on the values of the independent factors. Classification algorithms generate a model from a training set of data that can subsequently be used to predict the category of the dependent variable in fresh or previously unknown data. The model is assessed based on its ability to predict the category of the dependent variable in the test set accurately. The ultimate objective of classification is to create a model that can predict the category of the dependent variable for fresh data with high accuracy.

Classification is a machine learning approach that uses input characteristics to divide data into multiple classes or categories. Classification algorithms may be used to identify individuals with a high or low risk of developing U_N based on their C-peptide or glucose profile when identifying a U_N profile using C-peptide or glucose profile. A dataset containing measurements of C-peptide or glucose profile, as well as U_N profile, would be necessary to utilize classification to identify U_N profile. The dataset would then be separated into two subsets, one for training and one for testing. The classification algorithm is trained using the training set, and its performance is evaluated using the test set.

Once trained on the training set, the classification algorithm may be used to predict U_N profile based on C-peptide or glucose

profile for new patients. Metrics such as accuracy, precision, recall, and F1 score can be used to assess the classification algorithm's performance. These metrics indicate how successfully the algorithm classifies individuals as having a high or low risk of developing U_N depending on their C-peptide or glucose profile. In clinical practice, a classification system based on C-peptide or glucose profile might be used to identify individuals at high risk of developing U_N , allowing for early management and avoidance of illness.

A deeper knowledge in machine learning needs to be acquired to distinguish the types of analysis available and determine the appropriate analysis of the available data and the desired output. However, the readings and data extraction provide a clear picture of the most recent analytical alternatives accessible that are pertinent to this study regarding the U_N to some extent. The necessity for labeled training data is the key difference between supervised and unsupervised learning. Unsupervised machine learning uses unlabeled or raw data, whereas supervised machine learning uses labeled input and output training data. It can be observed that all 30 articles related to this study use supervised learning as an analysis tool. The purpose of supervised learning is to create a compact model of class label distribution in terms of predictor features.

Under the supervised learning approach, when the values of the predictor characteristics are known, but the value of the class label is unknown, the resulting classifier is used to assign class labels to the testing cases [53]. As supervised learning, classification and regression approaches are among the data analyses that are considered relevant for this study. These approaches were utilized to determine the adequacy and accuracy of the outcomes in this study. However, most of these 30 articles are utilizing the classification approach in acquiring the desired outcome due to the fact that most of the studies rely on identifying U_N based on analyzing the SI value.

The task of classification is to predict a discrete class label using a classification algorithm that predicts a continuous value in the form of a probability for a class label. Fig. 3 shows the most commonly used classification techniques under supervised learning approach. Classification algorithms used in machine learning utilize input training data for the purpose of predicting the likelihood or probability that the data that follows will fall into one of the predetermined categories. Several classification techniques can be classed into logic-based techniques, instance-based learning techniques, perceptron-based techniques, statistical learning techniques and support vector machine. Each class has a separate set of techniques as depicted in Fig. 3. Decision Tree (DT)[12, 13, 16, 17, 20, 22, 24, 26, 54], Support Vector Machine (SVM)[12-15, 20, 26], Bayesian Network [13, 15], K-Nearest Neighbor (KNN) [12, 19, 26, 28] and Multilayer Perceptron (MLP) [12, 14-17, 21, 55] were identified as the current classifiers within the 30 articles chosen, and these classifiers have been utilized often in studies over the previous 5 years. However, there are also techniques such as Logistic Regression (LR) [12, 14-16, 24, 27], Extreme Learning Machine Method (ELM) [18], eXtremeGradient Boosting (XG-Boost) [21, 22] and RF [12, 21, 22, 26, 27, 41, 51] that are not included in Fig. 3 used in the studies.

According to the analysis from the overall of the 30 studies, the top 4 techniques most used are DT, RF, SVM and LR. Nine

(9) out of 30 studies use the DT technique [12, 13, 16, 17, 20, 22, 24, 26, 54] whereas RF technique was used in 7 studies [12, 21, 22, 26, 27, 41, 51]. 6 studies each utilized the SVM [12-15, 20, 26] and LR [12, 14-16, 24, 27] technique. However, when considering the 3 categories of research gaps as depicted in Fig. 2, techniques were chosen based on the complexity of the analysis in order to meet the desired outcome. As portrayed in Fig. 4, it can be observed that the technique used in insulin secretion using deconvolution of plasma C-peptide concentrations study is RF. In addition, MLP and RF are the preferable techniques chosen to analyze the study of insulin secretion without taking into consideration or in-depth study of U_1 and U_2 . Furthermore, DT technique is the preferred technique to be employed when analyzing insulin sensitivity without identifying the insulin secretion.

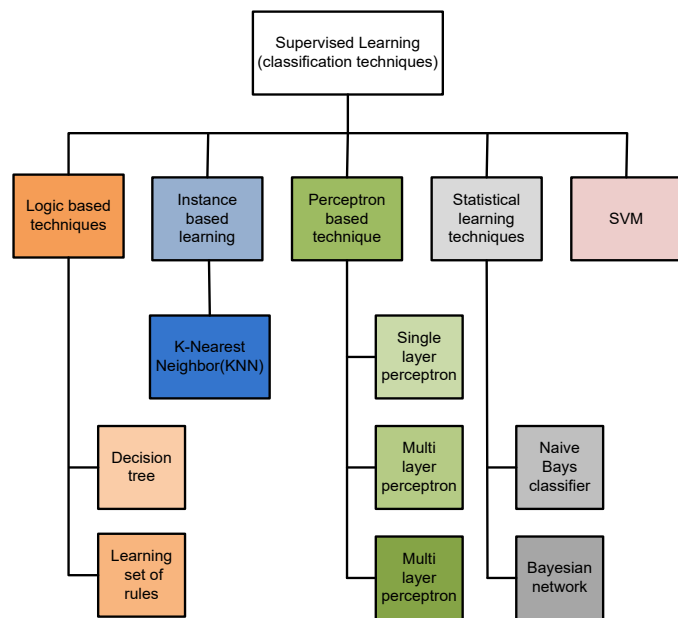


Fig. 3. Supervised learning (classification techniques)

Selection of data proves to be crucial in determining which technique is best suited for each finding. In the analysis of insulin sensitivity, the input data is merely physiological data such as anatomical function and also the metabolic state of the participants. In theory, DT is used to categorize or make predictions based on research objectives. Thus, with DT, the model is trained and tested on a data set that contains the desired categorization.

For the analysis of insulin secretion models, DT has a number of benefits, including interpretability, the capacity to manage non-linear connections, the ability to handle missing data, feature selection, robustness, and ease of implementation. DT are quite interpretable, which makes it simple for researchers to comprehend the connection between insulin secretion and the many contributing components, such as blood glucose levels, eating habits, and physical activity. Since the relationship between U_N and other parameters is frequently non-linear, DT is useful in modeling complicated and non-linear correlations between variables in analyses of U_N [56]. Moreover, DT can deal with missing data, which is frequently a problem in analyses of U_N models since it can come from a variety of

sources, including forgotten meals, missed insulin injections, or data collection mistakes [57, 58].

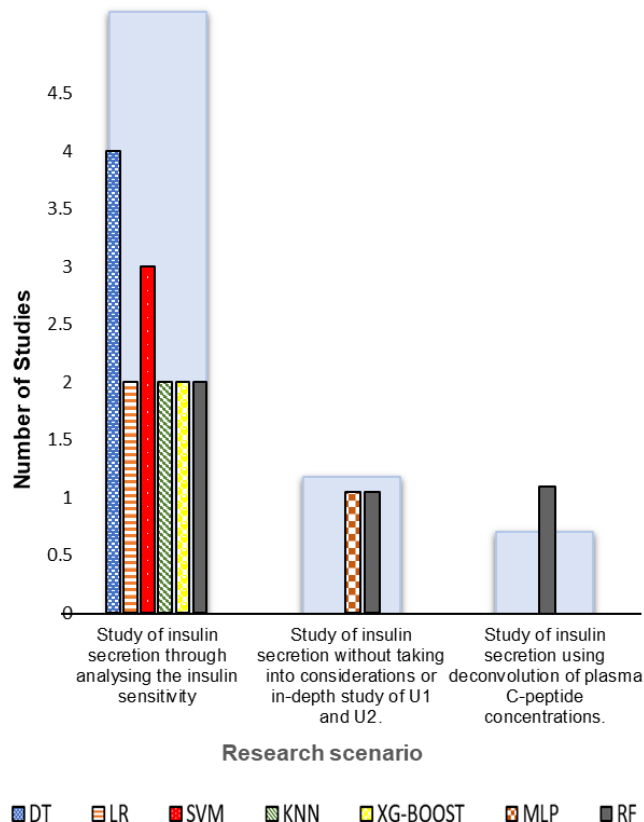


Fig. 4. The Most Chosen Machine Learning-Based Strategies in U_N Study

The most significant determinants affecting insulin secretion can be found using DT, which is helpful in creating focused interventions to increase U_N and regulate blood glucose levels in patients [59, 60]. Since the relationship between U_N and other factors may not be fully known, DT are well suited for U_N model analysis because they are relatively resilient to noisy and inconsistent data. DT are accessible to academics and physicians who may not have substantial machine learning knowledge due to their simplicity of implementation.

For the investigation of U_N models, RF have a number of benefits, such as increased accuracy, the capacity to handle high-dimensional data, robustness to noisy data, feature importance determination, and interpretability [61]. These benefits make RF a viable tool for enhancing our knowledge of insulin production and creating successful therapies to enhance glucose control in diabetic patients [62]. By merging the predictions of many DT, RF can increase the precision of U_N model analysis. This can lessen the overfitting and volatility that come with using a single DT.

In analyses of U_N models, where numerous factors might affect U_N , high-dimensional data can be handled by RF. Due to the intricate and unpredictable nature of U_N , RF are more resilient to noisy and inconsistent data, which is typical in U_N model research [63]. In order to create tailored interventions to increase U_N , it is possible to employ RF to assess the relative significance of several parameters in predicting U_N . Even though RF can be more complicated than individual decision

trees, they nonetheless offer some interpretability by enabling researchers to determine which elements are most strongly influencing the prediction.

Despite the advantages and disadvantages of these 2 techniques in identifying the participants specific U_N profile, there should be a thorough study on utilizing these supervised machine learning techniques (DT or RF) in identifying the participants specific U_N profile based on 2 critical inputs; 1) glucose excursion and 2) anatomical function (age, sex and body mass index). It is known that the regulation of BG by U_N is controlled by a physiological close-loop feedback-control system [64]. Thus, by employing the BG profile as an input, hypothetically it would provide a better identification of U_N profile. Additionally, by characterizing the U_1 and U_2 as a function of participants' anatomical characteristics, it would further improve the estimation of U_N profile.

IV. CONCLUSION

On five bibliographic databases, a total of 30 out of 97 papers that analyze the U_N using machine learning-based methods could be found. When quantifying the U_N profile in relation to the metabolic state of the individuals, DT and RF are the preferred options. More succinctly, it is understood that the available input data has a direct impact on the techniques that are chosen. Hypothetically, by combining all these input data, it would provide further insight into the pathogenesis of this diabetes. With the use of classification and regression techniques under supervised learning, further validation with larger and different cohorts as well as BG profile and anatomical function as inputs can further improve the early diagnosis capability. A successful outcome would provide valuable and direct information of pancreatic β -cells activity (U_N production) towards discriminating the metabolic status of an individual that leads to improved healthcare and wellness through individual specific therapies and proper management of diabetes.

V. ACKNOWLEDGMENT

Authors would like to thank Ministry of Education, Malaysia, for the Fundamental Research Funding grant (FRGS/1/2019/TK04/UITM/02/27) and Universiti Teknologi MARA, Cawangan Pulau Pinang for providing research facilities to run this study.

REFERENCES

- [1] WHO, "Insulin and associated devices: access for everybody: WHO stakeholder workshop, 21 and 23–25 September 2020," 2020.
- [2] N. Z. Arunah Chandran, "National Diabetes Registry Report 2020," Disease Control Division, Ministry of Health, Non Communicable Disease Section, 2021.
- [3] N. M. H. Osman *et al.*, "Effects of White Rice-Based Carbohydrates Diets on Body Weight and Metabolic Parameters in Rats," *Malaysian Journal Of Medicine and Health Sciences*, vol. 18, no. 1, pp. 234-240, 2022.
- [4] J. Kim, I. Jo, and H. Joung, "A rice-based traditional dietary pattern is associated with obesity in Korean adults," *Journal of the Academy of Nutrition and Dietetics*, vol. 112, no. 2, pp. 246-253, 2012.
- [5] E. A. Hu, A. Pan, V. Malik, and Q. Sun, "White rice consumption and risk of type 2 diabetes: meta-analysis and systematic review," *BMJ*, vol. 344, 2012.
- [6] V. S. Malik, B. M. Popkin, G. A. Bray, J.-P. Després, and F. B. Hu, "Sugar-sweetened beverages, obesity, type 2 diabetes mellitus, and cardiovascular disease risk," *Circulation*, vol. 121, no. 11, pp. 1356-1364, 2010.
- [7] J. E. Hall and M. E. Hall, *Guyton and Hall textbook of medical physiology e-Book*. Elsevier Health Sciences, 2020.
- [8] J. Ozougwu, K. Obimba, C. Belonwu, and C. Unakalamba, "The pathogenesis and pathophysiology of type 1 and type 2 diabetes mellitus," *J Physiol Pathophysiol*, vol. 4, no. 4, pp. 46-57, 2013.
- [9] N. A. Othman, "Analysis and optimisation of model-based insulin sensitivity and secretion tests," PhD Thesis, Faculty of Engineering, University of Canterbury, Christchurch, New Zealand, 2015.
- [10] E. Van Cauter, F. Mestrez, J. Sturis, and K. S. Polonsky, "Estimation of insulin secretion rates from C-peptide levels. Comparison of individual and standard kinetic parameters for C-peptide clearance," (in eng), *Diabetes*, vol. 41, no. 3, pp. 368-77, Mar 1992, doi: 10.2337/diab.41.3.368.
- [11] D. Tricò *et al.*, "Effects of Low-Carbohydrate versus Mediterranean Diets on Weight Loss, Glucose Metabolism, Insulin Kinetics and β -Cell Function in Morbidly Obese Individuals," *Nutrients*, vol. 13, no. 4, p. 1345, 2021.
- [12] R. Patil and S. Tamane, "A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Diabetes," *International Journal of Electrical and Computer Engineering*, vol. 8, pp. 3966-3975, 10/01 2018, doi: 10.11591/ijece.v8i5.pp3966-3975.
- [13] S. Kr, K. M. S. J, and K. S. Ravichandran, "A novel approach for finding diabetic mellitus using ensemble model for an optimized classification," *Asian Journal of Pharmaceutical and Clinical Research*, vol. 10, p. 15, 09/01 2017.
- [14] S. Kalagotla, S. Gangashetty, and K. Giridhar, "A Novel Stacking Technique for Prediction of Diabetes," *Computers in Biology and Medicine*, vol. 135, p. 104554, 06/01 2021, doi: 10.1016/j.compbiomed.2021.104554.
- [15] D. Cai, W. Liu, L. Ji, and D. Shi, "Bayesian optimization assisted meal bolus decision based on Gaussian processes learning and risk-sensitive control," *Control Engineering Practice*, vol. 114, p. 104881, 2021/09/01/ 2021.
- [16] S. Shamiluulu, B. Amirgaliyev, and L. Cherikbayeva, "Critical analysis of scikit-learn ML framework and weka ml toolbox over diabetes patients medical data," *News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences*, vol. 6, pp. 231-236, 01/01 2017.
- [17] M. Akman, M. K. Uçar, Z. Uçar, K. Uçar, B. Baraklı, and M. R. Bozkurt, "Determination of Body Fat Percentage by Gender Based with Photoplethysmography Signal Using Machine Learning Algorithm," *IRBM*, 2021/01/07/ 2021.
- [18] M. Shanthi, R. Marimuthu, S. N. Shivapriya, and R. Navaneethkrishnan, "Diagnosis of Diabetes using an Extreme Learning Machine Algorithm based Model," in *2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST)*, 23-25 Oct. 2019 2019, pp. 1-5, doi: 10.1109/ICAWS.2019.8923142.
- [19] E. Cordelli *et al.*, "Early experiences in 4D quantitative analysis of insulin granules in living beta-cells," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 3-6 Dec. 2018 2018, pp. 2009-2016, doi: 10.1109/BIBM.2018.8621293.
- [20] M. K. Uçar, Z. Uçar, F. Köksal, and N. Daldal, "Estimation of body fat percentage using hybrid machine learning algorithms," *Measurement*, vol. 167, p. 108173, 2021/01/01/ 2021.
- [21] S. Jangir, N. Joshi, M. Kumar, D. Choubey, S. Singh, and M. Verma, "Functional Link Convolutional Neural Network for the Classification of Diabetes Mellitus," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 37, 05/01 2021, doi: 10.1002/cnm.3496.
- [22] A. Agarwal, A. Sahoo, I. Das, S. Rautaray, and M. Pandey, "Homo Sapiens Diabetes Mellitus Detection and Classification," 2021, pp. 445-454.
- [23] I. Miguel-Escalada *et al.*, "Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes," *Nature Genetics*, vol. 51, no. 7, pp. 1137-1148, 2019/07/01 2019.
- [24] G. Koren, G. Nordon, K. Radinsky, and V. Shalev, "Identification of repurposable drugs with beneficial effects on glucose control in type 2 diabetes using machine learning," (in English), *Pharmacology Research & Perspectives*, Article vol. 7, no. 6, p. 5, Dec 2019, Art no. e529, doi: 10.1002/prp2.529.
- [25] Z. Lin *et al.*, "Machine Learning to Identify Metabolic Subtypes of Obesity: A Multi-Center Study," (in English), *Frontiers in*

- Endocrinology*, Original Research vol. 12, 2021-July-14 2021, doi: 10.3389/fendo.2021.713592.
- [26] M. Challa and D. C. R., "Optimized Machine Learning Approach for the Prediction of Diabetes-Mellitus," 2020, pp. 321-328.
- [27] W. Lu *et al.*, "Discovery of metabolic biomarkers for gestational diabetes mellitus in a Chinese population," *Nutrition & Metabolism*, vol. 18, no. 1, pp. 1-16, 2021.
- [28] C. Krishna and T. Sasikala, "Home based healthcare monitoring system for diabetes patients using IoT," in *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018*, 2019: Springer, pp. 676-686.
- [29] A. L. Lynam *et al.*, "Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults," *Diagnostic and Prognostic Research*, vol. 4, no. 1, p. 6, 2020/06/04 2020, doi: 10.1186/s41512-020-00075-2.
- [30] A. K. Dwivedi, "Analysis of computational intelligence techniques for diabetes mellitus prediction," *Neural Computing and Applications*, vol. 30, no. 12, pp. 3837-3845, 2018/12/01 2018, doi: 10.1007/s00521-017-2969-9.
- [31] M. Heydari, M. Teimouri, Z. Heshmati, and S. M. Alavinia, "Comparison of various classification algorithms in the diagnosis of type 2 diabetes in Iran," *International Journal of Diabetes in Developing Countries*, vol. 36, no. 2, pp. 167-173, 2016/06/01 2016, doi: 10.1007/s13410-015-0374-4.
- [32] M. Panwar, A. Acharyya, R. A. Shafik, and D. Biswas, "K-nearest neighbor based methodology for accurate diagnosis of diabetes mellitus," in *2016 Sixth International Symposium on Embedded Computing and System Design (ISED)*, 15-17 Dec. 2016 2016, pp. 132-136, doi: 10.1109/ISED.2016.7977069.
- [33] R. Bellazzi and A. Abu-Hanna, "Data mining technologies for blood glucose and diabetes management," (in eng), *J Diabetes Sci Technol*, vol. 3, no. 3, pp. 603-612, 2009, doi: 10.1177/193229680900300326.
- [34] M. R. Devi and J. M. Shylla, "Analysis of various data mining techniques to predict diabetes mellitus," *International journal of applied engineering research*, vol. 11, no. 1, pp. 727-730, 2016.
- [35] A. Z. Woldaregay, I. K. Launonen, E. Årsand, D. Albers, A. Holubová, and G. Hartvigsen, "Toward detecting infection incidence in people with type 1 diabetes using self-recorded data (Part 1): a novel framework for a personalized digital infectious disease detection system," *Journal of medical Internet research*, vol. 22, no. 8, p. e18911, 2020.
- [36] J. L. Dickson, J. Alswelder, C. A. Gunn, C. G. Pretty, and J. G. Chase, "A C-Peptide-Based Model of Pancreatic Insulin Secretion in Extremely Preterm Neonates in Intensive Care," (in eng), *J Diabetes Sci Technol*, vol. 10, no. 1, pp. 111-118, 2015, doi: 10.1177/1932296815596175.
- [37] G. Song, G. Pacini, B. Ahren, and D. Z. D'Argenio, "Glucagon increases insulin levels by stimulating insulin secretion without effect on insulin clearance in mice," (in English), *Peptides*, Article vol. 88, pp. 74-79, Feb 2017, doi: 10.1016/j.peptides.2016.12.012.
- [38] H. Frikke-Schmidt, P. Arvan, R. J. Seeley, and C. Cras-Meneur, "Improved in vivo imaging method for individual islets across the mouse pancreas reveals a heterogeneous insulin secretion response to glucose," (in English), *Scientific Reports*, Article vol. 11, no. 1, p. 15, Jan 2021, Art no. 603, doi: 10.1038/s41598-020-79727-8.
- [39] L. R. Cataldo *et al.*, "Prolonged Activation of the Htr2b Serotonin Receptor Impairs Glucose Stimulated Insulin Secretion and Mitochondrial Function in MIN6 Cells," (in English), *PLoS One*, Article vol. 12, no. 1, p. 16, Jan 2017, Art no. e0170213, doi: 10.1371/journal.pone.0170213.
- [40] S. Perveen, M. Shahbaz, M. S. Ansari, K. Keshavjee, and A. Guergachi, "A Hybrid Approach for Modeling Type 2 Diabetes Mellitus Progression," (in English), *Frontiers in Genetics*, Original Research vol. 10, no. 1076, 2020-January-07 2020, doi: 10.3389/fgene.2019.01076.
- [41] M. H. Lim, W. H. Lee, B. Jeon, and S. Kim, "A Blood Glucose Control Framework Based on Reinforcement Learning With Safety and Interpretability: In Silico Validation," *IEEE Access*, vol. 9, pp. 105756-105775, 2021, doi: 10.1109/ACCESS.2021.3100007.
- [42] W. S. Lee, J. Jo, and T. Song, "Machine learning for the diagnosis of early-stage diabetes using temporal glucose profiles," *Journal of the Korean Physical Society*, vol. 78, no. 5, pp. 373-378, 2021/03/01 2021, doi: 10.1007/s40042-021-00056-8.
- [43] M. Shifrin and H. Siegelmann, "Near-optimal insulin treatment for diabetes patients: A machine learning approach," *Artificial Intelligence in Medicine*, vol. 107, p. 101917, 2020/07/01/ 2020.
- [44] J. Hau, "Animal Models for Human Diseases," in *Sourcebook of Models for Biomedical Research*, P. M. Conn Ed. Totowa, NJ: Humana Press, 2008, pp. 3-8.
- [45] C. S. Levetan and S. M. Pierce, "Distinctions Between the Islets of Mice and Men: Implications for New Therapies for Type 1 and 2 Diabetes," *Endocrine Practice*, vol. 19, no. 2, pp. 301-312, 2013/03/01/ 2013.
- [46] T. B. Aradóttir, Z. Mahmoudi, H. Bengtsson, M. Jensen, D. Boiroux, and N. Poulsen, "An Adaptive Model-Based Approach to Personalized Basal Insulin Initiation in Type 2 Diabetes," 2019.
- [47] M. G. Pedersen, A. Corradin, G. M. Toffolo, and C. Cobelli, "A subcellular model of glucose-stimulated pancreatic insulin secretion," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 366, no. 1880, pp. 3525-3543, 2008.
- [48] J.-D. Lin *et al.*, "The first and second phase of insulin secretion in naive Chinese type 2 diabetes mellitus," *Metabolism - Clinical and Experimental*, vol. 59, no. 6, pp. 780-786, 2010, doi: 10.1016/j.metabol.2009.09.024.
- [49] P. D. Docherty *et al.*, "A spectrum of dynamic insulin sensitivity test protocols," *J Diabetes Sci Technol*, vol. 5, no. 6, pp. 1499-1508, 2011.
- [50] T. F. Lotz, "High resolution clinical model-based assessment of insulin sensitivity," PhD Thesis, Faculty of Engineering, University of Catenbury, Christchurch, New Zealand, 2007.
- [51] Y. Nunez Lopez, R. Retnakaran, B. Zinman, R. Pratley, and A. Seyhan, "Predicting and understanding the response to short-term intensive insulin therapy in people with early type 2 diabetes," *Molecular Metabolism*, vol. 20, 11/16 2018, doi: 10.1016/j.molmet.2018.11.003.
- [52] H. Pedersen, V. Gudmundsdóttir, M. Pedersen, C. Brorsson, S. Brunak, and R. Gupta, "Ranking factors involved in diabetes remission after bariatric surgery using machine-learning integrating clinical and genomic biomarkers. npj Genom Med. 2016; 1," ed, 2016.
- [53] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, no. 3, pp. 159-190, 2006/11/01 2006, doi: 10.1007/s10462-007-9052-3.
- [54] S. R. Khan *et al.*, "The discovery of novel predictive biomarkers and early-stage pathophysiology for the transition from gestational diabetes to type 2 diabetes," *Diabetologia*, vol. 62, pp. 687-703, 2019.
- [55] W. S. Lee, J. Jo, and T. Song, "Machine learning for the diagnosis of early-stage diabetes using temporal glucose profiles," *Journal of the Korean Physical Society*, vol. 78, pp. 373-378, 2021.
- [56] B. Åhrén and G. Pacini, "Importance of quantifying insulin secretion in relation to insulin sensitivity to accurately assess beta cell function in clinical studies," *European Journal of Endocrinology*, vol. 150, no. 2, pp. 97-104, 2004, doi: 10.1530/eje.0.1500097.
- [57] W. Z. Liu, A. P. White, S. G. Thompson, and M. A. Bramer, "Techniques for dealing with missing values in classification," in *Advances in Intelligent Data Analysis Reasoning about Data: Second International Symposium, IDA-97 London, UK, August 4-6, 1997 Proceedings 2*, 1997: Springer, pp. 527-536.
- [58] S. Gavankar and S. Sawarkar, "Decision tree: Review of techniques for missing values at training, testing and compatibility," in *2015 3rd international conference on artificial intelligence, modelling and simulation (AIMS)*, 2015: IEEE, pp. 122-126.
- [59] R. Patil and S. Tamane, "A comparative analysis on the evaluation of classification algorithms in the prediction of diabetes," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 5, pp. 3966-3975, 2018.
- [60] K. Donsa, S. Spat, P. Beck, T. R. Pieber, and A. Holzinger, "Towards personalization of diabetes therapy using computerized decision support and machine learning: some open problems and challenges," *Smart Health: Open Problems and Future Challenges*, pp. 237-260, 2015.
- [61] S. DuBrava *et al.*, "Using Random Forest Models to Identify Correlates of a Diabetic Peripheral Neuropathy Diagnosis from Electronic Health Record Data," *Pain Medicine*, vol. 18, no. 1, pp. 107-115, 2017, doi: 10.1093/pm/pnw096.
- [62] R. Casanova, S. Saldana, E. Y. Chew, R. P. Danis, C. M. Greven, and W. T. Ambrosius, "Application of Random Forests Methods to Diabetic Retinopathy Classification Analyses," *PLoS One*, vol. 9, no. 6, p. e98587, 2014, doi: 10.1371/journal.pone.0098587.
- [63] B. Sudharsan, M. Peebles, and M. Shomali, "Hypoglycemia prediction using machine learning models for patients with type 2 diabetes," *J Diabetes Sci Technol*, vol. 9, no. 1, pp. 86-90, 2014.
- [64] C. Postic *et al.*, "Dual roles for glucokinase in glucose homeostasis as determined by liver and pancreatic β cell-specific gene knock-outs using

Cre recombinase," *Journal of Biological Chemistry*, vol. 274, no. 1, pp. 305-315, 1999.